Research Proposal for Speech and Video Compression
(Proprietary)

Seenu S. Reddi
ReddiSS@AOL.COM
7/31/2002

Objective

Present day speech and video compression technologies are based on the state of the art that existed two decades ago (1980 – 1990 era). Though they do an adequate job of compression and streaming over the internet, the faster hardware presently available and recent algorithm breakthroughs make it possible to achieve a higher quality speech and images at lower bit rate coding. Also the prevalent technologies like MPEG do not make it possible to trade/use hardware speed for quality. An objective of this proposal is to achieve classical music quality and pitch free speech compression at 8 KBPS and a 10K pixel video stream at 100 KBPS. This is achievable by development of new compression methodologies and hardware with a consequent revolutionary impact on streaming media.

Approaches

The primary approach will be based on the noisization technique introduced in the web pages of the open source 16 KBPS speech compression for Linux and other platforms (www.rspq.org). The usual approach to speech/video compression is to remove the redundancies (or correlation) present in the virgin signal and code the residual. The problem has been that the residual is not completely white and still correlated. For instance, the residual still carries pitch spikes in speech and border information in images. (The latter is also true for interframe residuals). My personal efforts over a span of almost twenty years have been to solve the problem of devising strategies to classify the residual signal and fit this signal with artificially and computer generated specimens for a better, noise free reproduction of the original signal. Also the availability of memory and computational speed should be exploitable in these strategies. In technologies like MPEG, ad hoc (though ingenious) techniques have been used but they are, in my view, unattractive because of a lack of general structure and a powerful basis. (In a sense the approach is "Chinese" in the sense of using thousands of pictograms for writing instead of Indo-Greek-European limited character sets). However the noisization technique provides a simple approach to converting the different kinds of residuals and residual shapes to a standardized "specimens" with Gaussian (I think it is Gaussian – though it could be exponential) distributions. Once we achieve this we can use vector quantization and Shannon's coding principles to achieve the representation of residuals in a systematized and (near) optimal manner instead of the existing ad hoc practices.

Of course the noisization approach is computationally expensive but affords a way to utilize the speeds and memory of the computers for compression to low bit rates. The web site www.rspq.org gives open source code for compressing speech to 16 K to 18 K bit rates using noisization. (In my opinion the sound quality is better and fuller than that of MP3). A search of 64 K codewords is used but for better qualities and bit rates, one should venture into millions (2 ^ 20 ~ 2 ^ 24 or millions of codewords) and eventually into trillions (2 ^30 and higher). The research effort will focus on realization of this search via hardware (VLSI, USB/Firewire) and algorithmic efforts (fast and optimal/suboptimal generation of Gaussian "specimens", statistical aspects of noisization problems – see the open letter in www.rspq.org). The eventual goal is a realization of a high quality 8 KBPS speech compression algorithm suitable for streaming high quality audio over the internet.

The noisization approach is also applicable to video compression. The enclosed figure (for an online BMP image at www.rspq.org/rosh.bmp) represents the results of a preliminary application of this principle to a still image. The 256 X 256 pixel image ROSH is divided into 128 X 128 segments. Instead of the suboptimal DCT technique, a two dimensional 8 X 8 to 20 X 20 prediction is used to find the residual signal. Though one could compute the prediction coefficients using something like Norman-Levinson inversion techniques, I used a simple Widrow's LMS approach to compute the coefficients for the 128 X 128 segments and amazingly as the figure shows it is possible to eliminate most of the two dimensional redundancies and reduce the variance to 25 % of its original values. Of course it is always dangerous to generalize from one scenario (especially in an image or speech processing application). The research effort will be directed at finding the prediction coefficients using the lattice techniques developed by Itakura for speech but may have to be specialized for two dimensional video signals. Incidentally as far as I know this is one of the first instances of applying 64 to 256 two-dimensional predictive coefficients to get a residual image.

The figure also shows the effect of noisization. Whereas the histogram of the original image is scattered, the histogram of the reduced image shows a distribution with a flat top. The noisization makes the histogram more Gaussian (it could be exponential). The next process is to use "specimens" with this kind of distribution to represent/code the video residual.

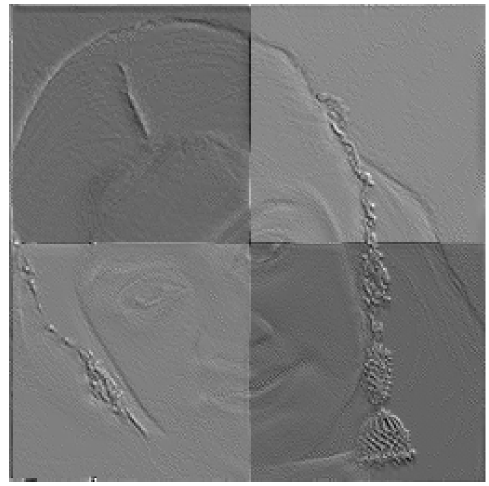
The noisization process has a great advantage in coding interframe differentials. The present day techniques attempt (with limited success) to predict frame to frame movements and use this information for coding. With noisization, no such steps are needed as it will "smear" the interframe differentials to uniform noise like salt and pepper images which can then be represented using reduced bit coding.

The final goal of the research efforts will be to develop a 128 KBPS (or less) video compression scheme for a 10 K pixel frame (352 X 288 CIF) at 15 fps. This implies a 0.5 ~ 0.75 bit per pixel and it appears in view of the temporal frame redundancies a feasible goal. But this statement does not minimize the efforts needed to achieve the same.

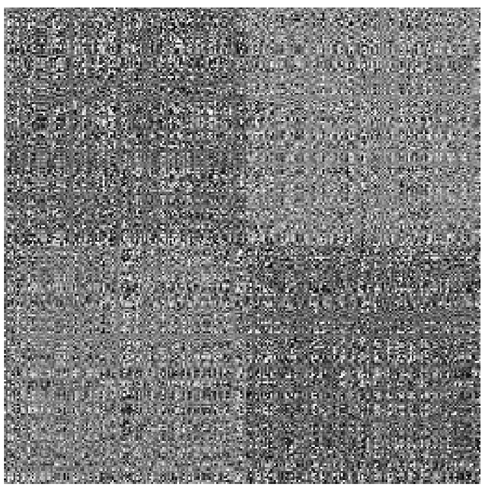
Goals

Realization of better and efficient video and speech compression and streaming technologies for internet and multimedia applications. Patents and publications for the ensuing technologies.

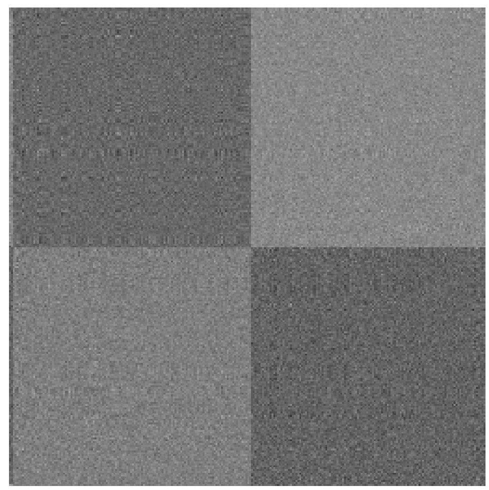
Original Image



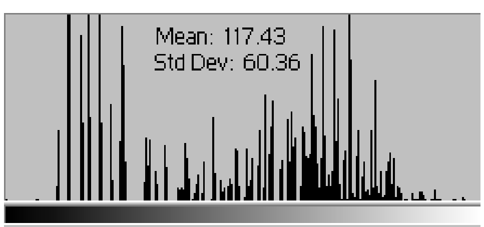
Error Image



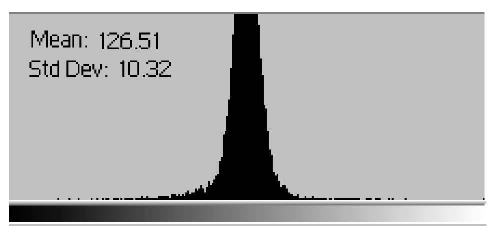
Noisized Original Image



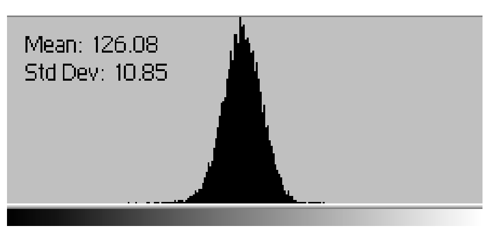
Noisized Error Image



Histogram of a Segment of the Original Image



Histogram of a Segment of the Error Image



Histogram of a Segment of the Noisized Error Image